Procesamiento del Lenguaje Natural, núm. 43 (2009), pp. 177-183 recibido 1-05-2009; aceptado 5-06-2009

Aplicación de la realimentación por relevancia con base textual para sistemas de recuperación visual *

Text-based relevance-feedback for content-based image retrieval systems

Arturo Montejo Ráez **José Manuel Perea Ortega**
Manuel Carlos Díaz Galiano **L. Alfonso Ureña López**

Departamento de Informática, Escuela Politécnica Superior
Universidad de Jaén, E-23071 - Jaén
{amontejo,jmperea,mcdiaz,laurena}@ujaen.es

Resumen: La realimentación por pseudo-relevancia es una técnica de uso habitual en sistemas de recuperación de información. Es sabido que los sistemas de recuperación basados en contenido (CBIR), como los de recuperación visual que permiten imágenes como objetos de consulta, adolecen de baja precisión. Esto es debido a la distancia entre la semántica de la imagen y sus características visuales. Es por ello que estos sistemas se ven beneficiados de un sistema de recuperación textual paralelo cuando dichas imágenes disponen de textos asociados. Se presentan diversos experimentos sobre realimentación sobre un sistema CBIR a partir de las listas textuales recuperadas. Los resultados obtenidos indican que estas técnicas pueden mejorar los métodos de fusión tradicionales.
Palabras clave: Recuperación de información, recuperación de imágenes basada en contenido

Abstract: Pseudo-relevance feedback is a common technique in information retrieval. Content Based Information Retrieval systems (CBIR), like visual retrieval engines, suffer from low performance due to the distance between the semantic interpretation of an image and its visual features. This is the reason why such a systems are empowered when textual information is associated to the images in the collection and the CBIR is combined with a textual retrieval system. In this paper, some experiments on applying text-based pseudo-relevance feedback on CBIR are reported. The results obtained show that these techniques may enhance the behavior of traditional fusion approaches.
Keywords: Information Retrieval, Content-Based Image Retrieval

1. Introducción

En sistemas de recuperación de información textuales, la realimentación por pseudo-relevancia ha probado ser una técnica valiosa para mejorar la relevancia de la lista de documentos obtenidos. Esta técnica consiste en obtener una primera lista de resultados a partir de la consulta original, para volver a lanzar una búsqueda con una consulta generada a partir de la original y los primeros documentos obtenidos.

Los sistemas de recuperación basados en contenido (CBIR) permite usar el mismo tipo de objetos a buscar como elementos de consulta. Un sistema CBIR de búsqueda de imágenes espera, de esta forma, un conjunto de imágenes de ejemplo para resolver la recuperación desde una colección visual (Datta et al., 2008). Existen ejemplos de sistemas que usan realimentación por relevancia en motores de recuperación visuales (Laaksonen et al., 2001; Müller et al., 2000). En cuando a la pseudo-realimentación sobre sistemas CBIR, son varias las propuestas realizadas (como por ejemplo (Torjmen, Pinel-Sauvagnat, y Boughanem, 2007)). si bien el trabajo de Ah-Pine et al. (Ah-Pine et al., 2009) ofrece una buena revisión de estas y otras técnicas asociadas a la recuperación combinada visual/textual.

Este trabajo se enmarca en la línea de investigación seguida por nuestro grupo, en sistemas de recuperación de información

* Esta investigación ha sido parcialmente financiada por la Junta de Andalucía, Consejería de Turismo y Deporte (FFIEXP06-TU2301-2007/000024)

(IR) híbridos, donde se utiliza información textual junto con información visual. Hemos participado varios años en las competiciones CLEF tanto en recuperación de imágenes genéricas (Cumbreras et al., 2007a; M.C. Díaz-Galiano, 2007a; Martín-Valdivia et al., 2005) como en recuperación de imágenes biomédicas (Díaz-Galiano et al., 2008; Díaz-Galiano et al., 2008; Díaz-Galiano et al., 2006), obteniendo en este último resultados prometedores. Además hemos realizado distintas experimentaciones utilizando expansión textual de la consulta (Díaz-Galiano, Martín-Valdivia, y Ureña-López, 2009), distintas heurísticas de traducción en colecciones multilingües y multimodales (Cumbreras et al., 2007b) o recuperación en otro tipo de colecciones multimodales (Díaz-Galiano et al., 2007; M.C. Díaz-Galiano, 2007b).

2. Descripción de la tarea

CLEF (Cross Language Evaluation Forum)[1] es un foro que aboga por el uso y desarrollo de aplicaciones para la gestión y manejo de librerías digitales. Para ello desarrollan infraestructuras de prueba, mejora y evaluación de sistemas de recuperación de información de diverso tipo. Dentro de las competiciones CLEF podemos encontrar la tarea ImageCLEF[2]. Esta tarea se encarga de evaluar distintos aspectos de los sistemas de recuperación de información multimodales, y más concretamente de aquellos que mezcla información visual y textual.

La tarea concreta sobre recuperación de imágenes médicas dentro del CLEF se conoce, genéricamente, como *ImageCLEFmed*. La colección de documentos proporcionada para esta subtarea hasta el 2006 (Müller et al., 2007) estaba formada por 4 subcolecciones de datos: CASImage, Pathopic, Peir y MIR, e incluyen unas 50.000 imágenes. Cada subcolección se organiza en *casos*. Un caso está formado por una o varias imágenes (dependiendo de la colección) y un conjunto de anotaciones en formato texto asociadas a dicha imagen. Las anotaciones están marcadas con etiquetas y constituyen los metadatos de la colección. Algunos casos incluyen más de una imagen relacionadas entre sí, por ejemplo, se puede tener una imagen de una radiografía de un fémur y,

[1] http://www.clef-campaign.org/

[2] http://www.imageclef.org/

```
<ID>3349</ID>
<Description>On the frontal and lateral
chest x-rays, perivascular haziness is
visible with a ground glass and diffuse
nodular infiltrate.
</Description>
<Diagnosis>Acute eosinophilic pneumonia
</Diagnosis>
<ClinicalPresentation> Patient with a
fever and respiratory insufficiency
since 5 days. </ClinicalPresentation>
<Commentary>The diagnosis was based on a
bronchoscopy with bronchoalveolar
lavage, demonstrating eosinophilia >
25%, as well as the absence of parasites
or any other pathogen.
...
```

Figura 1: Ejemplo parcial de un caso de la colección CASImage

asociada a esta imagen, disponer de otras que muestren secciones del mismo fémur, una resonancia magnética, una fotografía, etc. En la figura 1 podemos ver un ejemplo de la correspondencia entre un caso y sus imágenes.

Junto con la colección de imágenes y casos los organizadores de la tarea dieron un conjunto de consultas para evaluar los sistemas de recuperación. Dicho conjunto constaba de 30 consultas clasificadas como *visuales*, *textuales* o *mixtas*. Esta clasificación indica qué tipo de sistema de recuperación tiene más facilidades para encontrar mejores soluciones. Cada consulta está dividida en una parte textual traducida a 3 idiomas (inglés, alemán y francés) y una parte visual formada por una o varias imágenes. En la figura 2 se puede observar un ejemplo de una consulta de tipo visual.

3. Realimentación por relevancia

Nuestro método de realimentación por relevancia utiliza las dos listas de imágenes obtenidas mediante los dos procesos de recuperación descritos a continuación:

- **Recuperación de información textual**. Se recuperan imágenes a

```
<TOPIC>
<ID>27</ID>
<TYPE> visual </TYPE>
<EN-DESCRIPTION>
Show me images of a frontal head
MRI.
</EN-DESCRIPTION>
<DE-DESCRIPTION>
Zeige mir MR Frontalaufnahmen des
Kopfes.
</DE-DESCRIPTION>
<FR-DESCRIPTION>
Montre-moi des images IRM frontal du
crâne.
</FR-DESCRIPTION>
</TOPIC>
```

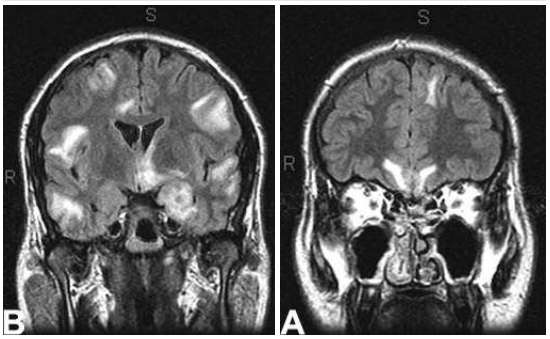

Figura 2: Ejemplo de una consulta de ImageCLEFmed

partir de un sistema IR tradicional como Lemur, haciendo uso de los textos asociados a las imágenes.

- **Recuperación de imágenes basada en contenido**. Se recuperan imágenes haciendo uso de un sistema CBIR como GIFT.

Una vez obtenidas ambas listas se genera una nueva consulta visual añadiendo imágenes adicionales. Estas imágenes adicionales pueden ser positivas (se desean imágenes similares) o negativas (ejemplos de imágenes que nada tienen que ver en la consulta). Para seleccionar las imágenes positivas y negativas se genera una lista a partir de las dos listas mencionadas anteriormente. En este proceso se calculan las posiciones de ranking para cada imagen en cada lista. A continuación, para las imágenes comunes a ambas listas exclusivamente, se calcula la diferencia de ranking obtenido para cada imagen:

$$Rank_{dif} = Rank_{textual} - Rank_{visual}$$

De esta forma, obtenemos la diferencia entre el ranking visual (obtenido del sistema CBIR) y el ranking textual (obtenido del sistema IR tradicional). Está comprobado el mejor comportamiento de los motores de recuperación textuales frente a los visuales en tareas de este tipo. Este es debido a que la semántica de la consulta queda mejor definida en el texto que en las características visuales de una imagen, ya que es habitual tener dos imágenes de características similares pero de semántica completamente diferente (por ejemplo, una radiografía de un brazo y una de una pierna).

Nuestra intención en la redefinición de la consulta visual es la de añadir imágenes semánticamente relevantes pero que añadan el mayor número de características visuales posible. Así pues, una imagen que aparece en las primeras posiciones de la lista textual pero en las últimas de la visual es candidata a ser un refuerzo positivo. En sentido inverso, una imagen en las primeras posiciones de la visual pero de las últimas en la textual es candidata a ser un refuerzo negativo. De esta manera, y usando la diferencia de ranking calculada, podemos tomar un número determinado de imágenes con mayor diferencia entre visual y textual como ejemplos positivos, y un número determinado de imágenes con diferencia negativa como ejemplos negativos.

Imágenes que estén altas en ambos rankings (visual y textual) y en posiciones similares obtendrán, por contra, valores de $Rank_{dif}$ menores, por lo que no serán consideradas en la reformulación de la consulta. Este efecto es el deseado, pues buscamos en la reformulación *añadir* imágenes con características visuales relevantes (no relevantes) en base a su relevancia (no relevancia) textual.

Esto contrasta con la forma tradicional de realimentación por pseudo-relevancia, donde siempre las imágenes en las primeras posiciones son las usadas para la generación de la nueva consulta. En este caso, y siguiendo una orientación de realimentación *intermedia* (Ah-Pine et al., 2009), buscamos mejorar la recuperación visual a partir de la relevancia más semántica del texto, frente a las características puramente visuales.

En nuestros experimentos pretendemos resolver dos cuestiones:

1. ¿Una realimentación visual basada en texto puede mejorar el comportamiento

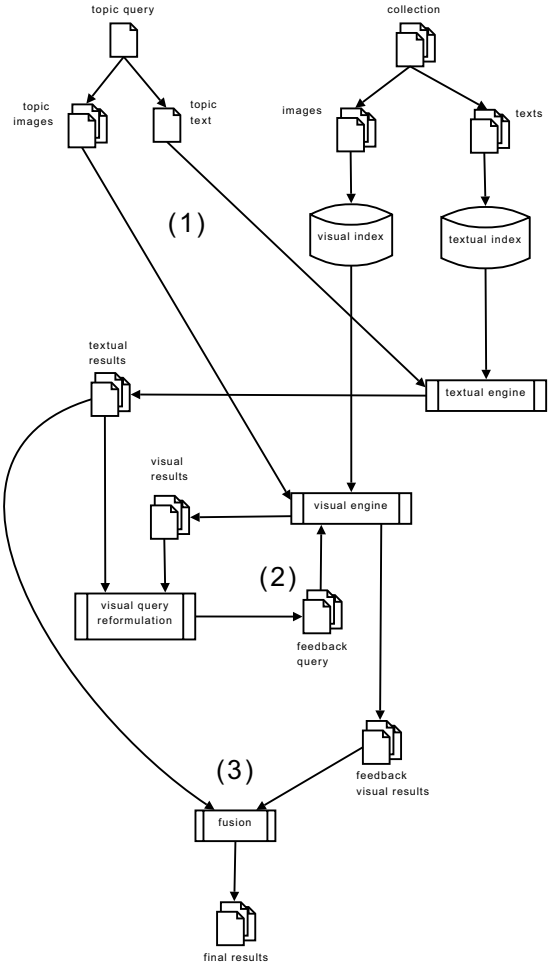

Figura 3: Proceso de obtención de la lista definitiva

del motor de recuperación visual significativamente?

2. De ser así, ¿el resultado de la fusión final también se verá mejorado o la mejora aportada por la recuperación textual ya queda patente en una fusión a partir de las lista originales, sin realimentación?

En la Figura 3 puede verse el proceso completo. En el paso (1) se obtienen las dos listas: visual y textual. En el paso (2) se regenera una consulta visual nueva, con refuerzos positivos y negativos, y se vuelve a consultar el índice visual. En el paso (3) se fusiona la lista visual resultante del proceso de realimentación con la lista textual para obtener una lista final.

4. Experimentos y resultados

En los experimentos llevados a cabo utilizamos dos parámetros fundamentales: el número de imágenes positivas (*npos*) y el número de imágenes negativas (*nneg*) a considerar en el proceso de realimentación. Durante los experimentos se han probado diferentes valores de *npos* y *nneg*: 0, 1, 2, 5, 10 y 50. Por otro lado, el número de resultados a devolver en ambas listas siempre se ha establecido en 1000, que es el número de resultados en la lista de referencia para su evaluación. Para la fusión, se ha realizado una combinación lineal basada en el RSV normalizado en el intervalo [0,1]. Esta combinación lineal responde a la fórmula:

$$RSV_f = RSV_t * \alpha + RSV_v * (1 - \alpha) \quad (1)$$

donde RSV_t es el *Retrieval Status Value* (valor de recuperación) obtenido con Lemur y RSV_v el obtenido con GIFT.

4.1. Caso base

Como caso base hemos considerado los resultados obtenidos aplicando cada uno de los dos subsistemas por separado, y la fusión de ambos sin realimentación. Como se puede observar en la Tabla 1, el resultado obtenido en valor de precisión media (*Mean Average Precision*, MAP) usando el sistema textual mejora con respecto al visual en 0.073 puntos. Por este motivo, para aplicar la realimentación, hemos considerado como imágenes de consulta negativas las que tienen un valor alto en el ranking visual y bajo en el textual y como imágenes positivas las que tienen un valor alto en el ranking textual y bajo en el visual.

Caso base	MAP
Sistema textual	0.1166
Sistema visual	0.0436
Fusion	0.1222

Tabla 1: Resultados sin realimentación (casos base)

Como puede verse, la fusión mejora los resultados obtenidos por cada motor de recuperación independientemente. Este valor de fusión ha sido obtenido con un pesado del 25 % para el ranking visual y del 75 % para el textual. Esto nos da a entender que, incluso si la recuperación visual dista mucho de tener un comportamiento deseable, sí que aporta algo de información que puede ser utilizada para aumentar la precisión de un motor combinado.

4.2. Utilizando todos los resultados de las listas

En una primera aproximación, utilizamos todos los valores devueltos por ambas listas para realizar los experimentos de realimentación por relevancia. Así, probamos con distintos valores de *npos* y *nneg*, obteniendo resultados no muy prometedores. Los mejores se muestran en la Tabla 2.

npos	*nneg*	MAP
0	1	0.0221
0	2	0.0256
0	5	0.0247
0	10	0.0244
0	50	0.0298
1	0	0.0101
2	0	0.0047

Tabla 2: Mejores resultados con realimentación usando todos los valores de las listas

La explicación de estos resultados se debe a la aleatoriedad de los últimos valores de ranking, pues ambos subsistemas son mediocres en cuanto a precisión. Es por ello que decidimos acometer una segunda aproximación en la que sólo tenemos en cuenta los n primeros resultados de ambas listas, intentando de este modo reducir dicha aleatoriedad.

4.3. Utilizando los n primeros resultados

En los experimentos utilizando los n primeros resultados de ambas listas, probamos con varios valores de n: 100, 1000, 5000 y 10000, obteniendo los mejores resultados con n=1000. Los mejores resultados se muestran en la Tabla 3.

n	*npos*	*nneg*	MAP
100	0	10	0.0463
100	10	10	0.0369
1000	1	0	0.0567
1000	2	0	0.0509
1000	5	0	**0.0580**
1000	10	0	0.0450
1000	50	0	0.0537
10000	0	10	0.0279
10000	0	50	0.0302

Tabla 3: Mejores resultados con retroalimentación usando los n primeros valores de las listas

Como se puede observar, el mejor resultado obtiene un MAP de 0.058, utilizando 5 imágenes de consulta positivas y ninguna negativa, devolviendo un total de 1000 imágenes. Esto supone una mejora de un 33 % con respecto al caso base en el que se utiliza el sistema visual (0.0436). En general, para $n = 1000$ se puede decir que los resultados con mezcla de imágenes positivas y negativas están todos por debajo de los que utilizan sólo imágenes positivas. Por otro lado, cuando $n = 100$, usando imágenes negativas los resultados se aproximan al caso base ($npos = 0$, $nneg = 0$).

4.4. Resultados finales con fusión

Nuestra evaluación no estaría completa sin los resultados obtenidos al fusionar la lista resultado de la realimentación con la lista textual, de tal manera que podamos comprobar si, a pesar de ser realimentado a partir del texto, la mejora neta del sistema con la fusión textual aumenta. La tabla 4 muestra los resultados de distintos valores lineales de fusión de listas realimentadas sólo con imágenes positivas (que son las que, en general, aportan mejores resultados). La primera línea correspondería con el caso base, pues no se usan imágenes adicionales en la reconsulta antes de la fusión. La mejora obtenida del mejor caso con respecto al caso base es de tan sólo un 4 %.

npos	*nneg*	*Vw*	*Tw*	MAP
0	0	0.25	0.75	*0.1222*
1	0	0.25	0.75	0.1229
5	0	0.50	0.50	0.1235
10	0	0.20	0.80	0.1242
10	0	0.25	0.75	0.1247
5	0	0.20	0.80	0.1264
5	0	0.25	0.75	0.1264
50	0	0.20	0.80	**0.1266**
50	0	0.25	0.75	**0.1266**

Tabla 4: Resultados finales de fusión tras realimentación

Vw y *Tw* indican el peso aplicado sobre los RSVs normalizados de la listas visual y textual respectivamente de tal manera que el RSV final es:

$$RSV = RSV_{visual} * Vw + RSV_{textual} * Tw \quad (2)$$

Como puede observarse en la tabla, los

valores de *Tw* son más altos que los de *Vw*. Esto es así porque sólo se muestran aquellos que mejor resultados han dado, siendo además consistentes con los pesados en fusión aplicados en otras tareas de recuperación visual (Díaz-Galiano et al., 2008; Díaz-Galiano et al., 2008; Díaz-Galiano et al., 2006).

5. Conclusiones

Podemos observar una mejora importante sobre la recuperación visual, por lo que la realimentación con base textual parece interesante en estos entornos. Además, esta mejora se muestra robusta combinada con otras técnicas como la fusión de listas visuales y textuales, proporcionando en última estancia una mejora general.

En la Tabla 4 encontramos un resultado que puede parecer anómalo: si bien la mejor realimentación visual se obtiene con 5 imágenes positivas, para la fusión hay, en cambio, una pequeña ventaja al usar 50 imágenes positivas. Esta diferencia no es significativa, pero nos lleva a pensar que existen otros factores que pueden influir en la mejora de la lista final. En concreto, esperamos aplicar técnicas de reordenación sobre los primeros resultados (Chen y Karger, 2006), de tal forma que se estudie el efecto de la diversidad de los resultados.

A la hora de establecer conclusiones al trabajo presentado es importante reconocer la debilidad estadística que pueden conllevar estos datos por varias razones: no se ha comprobado el comportamiento del método en otras colecciones. Es por ello que como trabajo futuro se impone comprobar este sistema sobre otras colecciones de objetos que combinen información visual y textual. Adicionalmente, experimentos aplicando realimentación por pseudo-relevancia sólo sobre texto e imágenes de forma separada son necesarios para poder contrastar con amplitud la capacidad del método propuesto.

Bibliografía

Ah-Pine, Julien, Marco Bressan, Stephane Clinchant, Gabriela Csurka, Yves Hoppenot, y Jean-Michel Renders. 2009. Crossing textual and visual content in different application scenarios. *Multimedia Tools Appl.*, 42(1):31–56.

Chen, Harr y David R. Karger. 2006. Less is more: probabilistic models for retrieving fewer relevant documents. En EfthimisÑ. Efthimiadis Susan T. Dumais David Hawking, y Kalervo Järvelin, editores, *SIGIR*, páginas 429–436. ACM.

Cumbreras, Miguel Angel García, Manuel Carlos Díaz-Galiano, Maria Teresa Martín-Valdivia, Arturo Montejo Ráez, y Luis Alfonso Ureña López. 2007a. Sinai system: Combining ir systems at imageclefphoto 2007. En Carol Peters Valentin Jijkoun Thomas Mandl Henning Müller Douglas W. Oard Anselmo Peñas Vivien Petras, y Diana Santos, editores, *CLEF*, volumen 5152 de *Lecture Notes in Computer Science*, páginas 512–517. Springer.

Cumbreras, Miguel Angel García, Maria Teresa Martín-Valdivia, Luis Alfonso Ureña López, Manuel Carlos Díaz-Galiano, y Arturo Montejo Ráez. 2007b. Using translation heuristics to improve a multimodal and multilingual information retrieval system. En Francesco Masulli Sushmita Mitra, y Gabriella Pasi, editores, *WILF*, volumen 4578 de *Lecture Notes in Computer Science*, páginas 438–446. Springer.

Datta, Ritendra, Dhiraj Joshi, Jia Li, y James Ze Wang. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2).

Díaz-Galiano, M. C., M. T. Martín-Valdivia, y L. A. Ureña-López. 2009. Query expansion with a medical ontology to improve a multimodal information retrieval system. *Computers in Biology and Medicine*, 39(4):396–403.

Díaz-Galiano, Manuel Carlos, Miguel Angel García Cumbreras, M. T. Martín-Valdivia, Arturo Montejo Ráez, y L. A. Ureña-López. 2008. Integrating mesh ontology to improve medical information retrieval. En Carol Peters Valentin Jijkoun Thomas Mandl Henning Müller Douglas W. Oard Anselmo Peñas Vivien Petras, y Diana Santos, editores, *Advances in Multilingual and Multimodal Information Retrieval*, volumen 5152 de *Lecture Notes in Computer Science*, páginas 601–606. Springer.

Díaz-Galiano, Manuel Carlos, Miguel Angel García Cumbreras, Maite Teresa

Martín-Valdivia, Arturo Montejo Ráez, y Luis Alfonso Ureña López. 2006. Using information gain to improve the imageclef 2006 collection. En Carol Peters Paul Clough Fredric C. Gey Jussi Karlgren Bernardo Magnini Douglas W. Oard Maarten de Rijke, y Maximilian Stempfhuber, editores, *CLEF*, volumen 4730 de *Lecture Notes in Computer Science*, páginas 711–714. Springer.

Díaz-Galiano, Manuel Carlos, Maria Teresa Martín-Valdivia, Miguel Angel García Cumbreras, y Luis Alfonso Ureña López. 2007. Using information gain to filter information in clef cl-sr track. En Carol Peters Valentin Jijkoun Thomas Mandl Henning Müller Douglas W. Oard Anselmo Peñas Vivien Petras, y Diana Santos, editores, *CLEF*, volumen 5152 de *Lecture Notes in Computer Science*, páginas 719–724. Springer.

Díaz-Galiano, M.C., M.A. García-Cumbreras, M.T. Martín-Valdivia, L.A. Ureña-López, y A. Montejo-Raez. 2008. Sinai at imageclefmed 2008.

Laaksonen, Jorma, Markus Koskela, Sami Laakso, y Erkki Oja. 2001. Self-organising maps as a relevance feedback technique in content-based image retrieval. *Pattern Anal. Appl.*, 4(2-3):140–152.

Martín-Valdivia, Maite Teresa, Miguel Angel García Cumbreras, Manuel Carlos Díaz-Galiano, Luis Alfonso Ureña López, y Arturo Montejo Ráez. 2005. The university of jaén at imageclef 2005: Adhoc and medical tasks. En Carol Peters Fredric C. Gey Julio Gonzalo Henning Müller Gareth J. F. Jones Michael Kluck Bernardo Magnini, y Maarten de Rijke, editores, *CLEF*, volumen 4022 de *Lecture Notes in Computer Science*, páginas 612–621. Springer.

M.C. Díaz-Galiano, M.A. García-Cumbreras, M.T. Martín-Valdivia A. Montejo-Raez L.A. Ureña-López. 2007a. Sinai at imageclef 2007.

M.C. Díaz-Galiano, J.M. Perea-Ortega, M.T. Martín-Valdivia A. Montejo-Ráez L.A. Ureña-López. 2007b. Sinai at trecvid 2007.

Müller, Henning, Wolfgang Müller 0002, StÃ©phane Marchand-Maillet, Thierry Pun, y David Squire. 2000. Strategies for positive and negative relevance feedback in image retrieval. En *ICPR*, páginas 5043–5042.

Müller, Henning, Thomas Deselaers, Thomas M. Lehmann, Paul D. Clough, y William Hersh. 2007. Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. En *CLEF Workshop 2006*, volumen 4730 de *LNCS*, páginas 595–608, Alicante, Spain, 20/09/2006. Springer, Springer.

Torjmen, Mouna, Karen Pinel-Sauvagnat, y Mohand Boughanem. 2007. Using pseudo-relevance feedback to improve image retrieval results. En Carol Peters Valentin Jijkoun Thomas Mandl Henning Müller Douglas W. Oard Anselmo Peñas Vivien Petras, y Diana Santos, editores, *CLEF*, volumen 5152 de *Lecture Notes in Computer Science*, páginas 665–673. Springer.